DOCUMENT RESUME

ED 190 232 PS 011 586

AUTHOR Hilliard, Asa G. III
TITLE The Ideology of Intelligence and I.Q. Magic in Education.
PUB DATE Sep 79
NOTE 12p.; Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Practices; Identification; *Intelligence; Intelligence Quotient; *Intelligence Tests; Prediction; Psychological Testing; Psychometrics; Selection; Test Bias; Testing Problems; *Test Interpretation; *Test Validity

ABSTRACT

The standardized IQ tests which are in use in the schools are scientifically and pedagogically without merit. The construct "intelligence" is a hypothetical notion whose valid expression has yet to be born. IQ tests and the construct of intelligence can be discarded at present, and teaching strategies would be unaffected. To successful teachers the tests are at best a sure nuisance and at worst a reactive influence on teaching and learning. The tests are not simply culturally biased. That bias is only a symptom of the problem which is their scientific inadequacy. To say that "they are the best we have," is not to say that they contribute anything useful at all to instruction. The construct "intelligence" is embryonic and has heuristic value for research. Its utility for instruction remains to be demonstrated. School teachers and students should be relieved of the burden of this bad science and psychological ideology. Testmakers should come again when this product can help to make education better. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED1902232

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE IDEOLOGY OF INTELLIGENCE

AND

I.Q. MAGIC IN EDUCATION

By

Asa G. Hilliard, III

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Asa G. Hilliard, III

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

(Symposium, "Is the Construct of Intelligence a 20th Century Myth," New York American Psychological Association Meeting, 1979)

"I don't see any point in I.Q. testing in schools. There's no more point in measuring I.Q. than in measuring the basal metabolic rate. Pupils, teachers, and parents need to know whether a pupil is learning what the school is trying to teach, but I can't see that they need to know the child's I.Q." (Arthur Jensen, 1979)

I believe that the future will bring increasingly valid models of how the mind works in the learning process. It may also bring pedagogical applications for these valid models, applications which improve teaching and learning. It is even possible that the testing of mental dynamics may be accomplished by standardized procedures which can be performed cheaply, efficiently, and for the entrepreneur, even at a profit. That time is not now for the standardized I.Q. tests which are widespread use in schools.

Let me be clear at the outset. I am looking primarily at the utility of both I.Q. tests and the construct intelligence for teaching practice. While the construct intelligence remains undefined operationally among the community of scholars, and while the tests to measure intelligent behavior cannot be shown to be universally valid instruments, these are technical matters for designers within the profession. During a research phase, wide lattitude can be tolerated. However, when professional applications are made and where legal mandates for the use of instruments are involved, rigor in validity must be assured.

No position can be taken on whether "it" ("intelligence") is genetic or environmental until we have sufficient data to determine if "it" even exists; and if so, in what form. Similarly, it is premature to debate whether sorting by intelligence should be done by tests for school purposes unless the existence of intelligence can first be

PS011586

established and can be validly related to instruction. The test for utility must be made, determining if thinking with the construct of intelligence and/or testing with I.Q. tests makes a positive impact on the teaching-learning environment.

Finally, let me emphasize that I do support valid assessment. I also believe that psychology has much to offer teaching. At present, I.Q. test offering is only patent medicine.

The construct "intelligence" and the I.Q. tests which were designed to measure the behavior implied by the construct were fabricated and were applied in education prior to the time that mental functions were even described clearly. Mental "measure" went from "research" through development to general application in an amazingly short time. Public school policy makers needed something like I.Q. tests; and presto! they were there. (Levine, 1976) "Intelligence" then, as now, was said by I.Q. test advocates to be measured precisely, before it could even be defined operationally in a common way by the community of scientists. There was not then, nor has there been since then, any general professional requirement that this undefined substance be measured in a uniform and rigorous way or that it be measured with instruments that yield comparable data. For example, the "subtests" on various I.Q. tests follow every conceivable pattern. Does each represent a component of intelligence? If not, then what is the meaning of a "subtest"? If so, does intelligence vary with the test?

Still, no matter how poor a construct, instrument, or procedure for mental measurement might appear to be, any serious educator could and would overlook the lack of construct and instrument refinement, if the use of constructs, instruments, or procedures resulted in improved performance in teaching and learning. Has this happened? I know of no data to show that it has. There are few professional researchers who seem willing to dare to ask the question.

It can be observed that the construct intelligence and the I.Q. tests which purport to measure the behavior implied by it are in almost universal use in the public schools. Among current popular uses for I.Q. tests are the following:

1. To determine a child's "readiness" for kindergarten.
2. To predict a person's future academic performance.
3. To classify a pupil for placement in special schools or programs.
4. To determine if a child is "socially competent."
5. To "diagnose" learning difficulties.

Yet we may ask again, is teaching and learning improved as a consequence of the use of I.Q. tests? The startling thing which one discovers when trying to answer that question, is that it is almost impossible even to establish criteria by which answers to the questions can be judged! It is really not clear just what is supposed to happen under ideal circumstances. Some of the reasons for this are simple. Professional language in testing is full of gross ambiguity. The full range of assumptions upon which professional discourse is based is seldom made explicit. Where implicit assumptions can be inferred, they are hopelessly confounded. This confounding is evident as discussants in the intelligence and I.Q. debate slide back and forth from one set of assumptions to another, giving little evidence that they are aware that the shifts have been made. Let's take an example. I.Q. testing may be used for a variety of school purposes. They may be used for sorting and classification, for diagnosis of learning difficulties, for the development of individual educational plans, for research on thinking, and for selection for admission to education opportunities. Yet it is anything but clear just how a given I.Q. test such as a Wechsler or a Binet can be used to serve all these diverse needs. If, for example, I.Q. test advocates are challenged to demonstrate "prescriptive" or pedagogical validity (Gallager, 1976) for the test for a particular purpose, arguments which more logically support an entirely different purpose may be marshalled by test defenders. The arguments in support of the validation of a particular test as an individual diagnostic device would hardly be expected to be the same as arguments in support of that same test as a program sorting device.

Therefore, for example, the diagnosis of many African-American children by the use of a standardized I.Q. test may be challenged as invalid because such tests use an unfamiliar European-American vocabulary as a part of the "measure" of "mental capacity." This challenge is met frequently by a spurious argument. The spurious argument is that "all Americans should, for practical reasons, master the general culture." Notice that the challenge raised questions regarding the valid measurement of such things on I.Q. tests as remembering large numbers of words, using words"properly," etc. If different children are to be compared, these things should be "measured" using a vocabulary which all tested children have had an equal opportunity to learn. I.Q. test advocates' responses to the challenge above tend to change the focus of the discussion from points about the "measurement" of "mental functions" to a focus on the practical utility of a common language. Such a response about utility is true but irrelevant to the issue of measuring the child's dynamic patterns of thought.

There are other dimensions to the discussion which create similar confounding in discourse, for example, each time the audience for information changes, the nature of the information which is needed changes as well. Policy makers might wish to know if I.Q. tests "work," if they are cheap, if they can identify "gifted children."

On the other hand, a teacher may wish for information on a special strategy to use on Monday with Johnny Jones. A prescriptive diagnosis and a policy recommendation probably require different information, or information at different levels of refinement. Yet the teacher and the policy maker are most likely to get the same information, a raw I.Q. score or scores or a gross label such as "EMR."

Current school uses of I.Q. tests tend to reflect an emphasis on "prediction" and/or "diagnosis." The I.Q. test is supposed to tell us what a child's future school performance will be and, by implication, what the limits of a child's performance must be. The I.Q. test is supposed to tell us what is "broken," "disabled," or "underdeveloped" in the child's thinking process. Both of these uses, diagnoses and prediction, offer an excellent opportunity to illuminate just how underdeveloped intelligence and I.Q. ideology and practice are.

Let's take them one at a time. I.Q. tests tend to correlate positively with other I.Q. tests and with some school grades under certain conditions. Yet other I.Q. tests and school grades both tend to use test items which are quite similar to those on the I.Q. test with which we begin. Therefore, we should not live in awe of the fact that a given thing tends to be mildly associated with something quite like itself. On the other hand, we should be embarrassed as scientists that virtually all of the I.Q. correlations which are reported in the literature are based on many studies that repeat a simple unscientific error. That unscientific error is the failure to control for known major sources of variation in experimental or general testing conditions, and to ignore this failure in subsequent interpretation. To be specific, there is an almost universal failure to control for instructional "treatment" in the validity studies which have been done. Studies using pre-measures on I.Q. tests and post-measures on school grades presume equivalence of "treatment" or teaching among subjects among comparison groups. Nothing could be further from the truth in most cases in the schools (Hamilton,Rist). Moreover, there are compelling data which suggest that if this were done, the famous or infamous I.Q. correlation would evaporate (Fuller, 1978) (Johntz, 1976) (Freire, 1973). This seems to be a taboo area for most I.Q. research!

But what about diagnosis? The use of the word "diagnosis" implies a knowledge of how thinking ought to work. When a professional in applied areas such as clinical psychology, school psychology, educational psychology, or school teaching uses the word "diagnosis," there is a further implication. That implication is that there is a systematic practice or pedagogy which if properly applied, will work to produce student achievement. In such a teaching model, such systematic pedagogy or "treatment" ought to be public professional knowledge and should be endorsed as valid by the profession. Medical doctors might call this "standard procedure." Without valid "standard

procedure," any I.Q. "diagnosis" for the applied professional is professionally meaningless and useless. Yet, there are in fact no "standard procedures" for teachers in places where I.Q. tests are in use. Theories of teaching do exist in abundance. Many teachers and researchers have described how some teachers function. But educators have yet to recognize or to sanction a set of common valid pedagies. Teacher education is still quite anarchic, and so is common teaching practice. This eliminates even the possibility of another critical matter. That matter is that there be a valid link between testing and pedagogy, between I.Q. testing and teaching strategies, between "diagnosis" and "prescription," and between both of these and "healing." Is there anyone here or elsewhere who is willing to stake their professional reputation on a claim that these links can be demonstrated?

Without valid or meaningful prediction, we are left with but one major use of I.Q. tests in education. That use is to sort students into categories so that they may be treated in special ways. This assumes that the classifications will yield intellectually homogenous groups of students who can and should be given a unique educational treatment as a group. Note again: this special unique treatment is mysterious. It is implied but never described. Sorting can be accomplished by the use of I.Q. tests. But the same sorting outcomes could also be accomplished almost as easily by use of the social class indices of family income, family educational level, and family prestige, and by use of skin color, (Nader, 1979). Any other arbitrary marker could be used for sorting to identify a part of the population which is to be excluded from normal opportunity. But this kind of sorting is clearly political, and not psychological or educational in any professional sense, (Hilliard, T., 1979). To be professional, the testing and pedagogy link would have to be validated.

Basically then the whole I.Q. test operation rides on three legs. They are:

1. Prediction
2. Diagnosis
3. Sorting

None of these as yet can be regarded as valid educational practice. There are no data to show that student performance is improved because of these three uses of tests.

The I.Q. test may serve well as a clinical interview protocol for a psychologist who is thoroughly familiar with a student's culture. Further I.Q. tests or items should be permitted for purposes of research for test development. I take no issue whatsoever with these

uses. It is only when as a clinical aid the I.Q. test is offered as a "measurement" device in a scientific sense, or when a research tool is passed off as a valid applied device that the illegitimate imposition of tests on clients must be questioned.

I hope that I have made it clear that it is not simply the misuse of currently used I.Q. tests but the inherent scientific inadequacy of the tests themselves that is being questioned. Further, I hope that it is also clear that I have made no special plea here for correcting the cultural bias of currently used I.Q. tests. The problem is far more grave than that. The cultural bias only shows us that standardized mass-produced "measurement" is impossible when variable cultural material is being aggregated in cross-cultural settings. This is aggregating apples and oranges. The culture and measurement issue is a matter of science first and then equity. Clearly, Pandora's box will be opened in the mental measurement laboratory on the very day that cultural anthropologists and socio-linguists are invited to look at what we do. No existing standardized I.Q. test can survive that kind of scientific scrutiny. The whole I.Q. testing movement reflects either an ignorance of or an unwillingness to deal with relevant academic data, especially socio-linguistic. The incomplete literature review in most research on the validation of I.Q. tests will reveal this scientific defect.

If Nero did indeed fiddle while Rome burned, then it is a fitting analogy for I.Q. test advocates and for those who fail to teach children successfully. While test advocates conduct their pseudo "measurements," there are numerous examples of outstanding pedagogy in America and in the world which proceed without them. There are exciting examples which include those where there is dramatic achievement for children who should not have been able to achieve so well, based upon their I.Q. scores, (Freire, 1973; Johntz, 1976; Fuller, 1978; Hilliard, 1979). It should be sobering to note that in my experience with teaching that succeeds, I do not know of a single instance in which the educators or psychologists relied upon I.Q. tests! The other side of that is this: I have yet to see a demonstration anywhere to show that the use of I.Q. tests make a positive difference in the achievement of children. Researchers have looked at every child, family, or social-class variable imaginable, yet the empirical proof of the pedagogical utility of I.Q. tests remains to be done.

In my experience of observing successful teaching and in reviewing the literature on that same subject, I have yet to come across teachers or psychologists who utilize the construct of intelligence directly in their work. They simply do not talk in terms of a measured amount of student capacity. I am aware that the systematic observation of learners has begun to help us to understand the teaching and learning process better. Piaget's work seems to have the potential for a growing application in teaching, (Elkind, 1971; Furth, 1977).

The systematic observation of learners over long periods of time and across broad cultural groups will eventually yield basic scientific knowledge, especially when the unique patterns of learners are observed, accounted for and interpreted a la Piaget. The spate of publications which deal with the application of Piaget to the classroom, though incomplete and sometimes controversial, may be compared favorably to the relative absence of widely used publications which spell out the use of "intelligence" and I.Q. in the classroom. What can a teacher do with "g?" even if it turns out to be more than an artifact of particular approaches to testing and data analysis. When "g" is "measured," what more does a practicing professional know about a child than he or she knew before? Few I.Q. technicians seem to have the courage to go beyond academic fortune telling at about the same level of specificity as our daily horoscopes in the Toonerville Chronicle. There simply is no significantly useful information in the test for teachers.

Intelligence as a construct and currently used I.Q. tests fail education, not merely because of their readily apparent technical poverty, or because of demonstrable cultural bias (Hilliard, 1979), but because they are, at present, useless as instructional tools.

The repair of the damage which has been done in the quest for I.Q. and intelligence can be made only if work is begun on the right problems! This requires that fundamental confusion be overcome.

1. Statistical bias must no longer be confused with cultural bias.

   To address the issue of cultural bias there must be a sophisticated identification of cultural groups, an understanding of culture and cognition, and an understanding of socio-linguistic principles. The fact that items may appear to "work the same way" within two different "cultural" groups when simple statistical calculations are used does not deal effectively with the cultural issue. How are the cultural groups to be identified? How does sample selection proceed?

2. English must no longer be confused with language.

   Alas, English is only a language. It is only one of many. Even at that it is a polyglot language, made up of a basically Germanic morphology and a basically Romance or Latin vocabulary. Just as with many other linguistic amalgams, it has utilitarian value. However, thinking can be expressed in every language. English does not own thinking. Getting English rules "right" is not necessarily the same as getting thinking right. Therefore, the deep structure of language (Chomsky, 1957) (Levi-Strauss, 1966) can no longer be confused with the surface structure (a particular language such as English). Standard or "correct" English should no longer be confused as a unique expression of thinking. The implications of this for standardized testing which use "standard English" are immense.

3. The aggregation of numbers (test scores) must no longer be confused with the aggregation of comparable "units of mental behavior."

   With what logic can a subject's response to a block design item be aggregated with a response to a vocabulary item? This is measurement??

4. Prediction must no longer be confused with diagnosis.

   The noting of a small association between two sets of scores (I.Q. and achievement) is not an explanation of the association.

5. Statistical categories must no longer be confused with behavioral functions.

   For example, a "gifted" person cannot be described simply as a person who "falls into the top 2%" of scorers on an I.Q. test. A description of the unique mental functions must be made. There is no reason whatsoever for the frequency of functions to appear in a population by prior definition rather than by actual experience.

6. Non-discriminatory assessment must no longer be confused with valid assessment.

   The search for "culture free" assessment had to be a failure almost by definition. Virtually all communicative human behaviors appear to be human creativities, or simply "cultural" material. Culture must be used in all assessment, but not the same culture in all assessment. Further "non-discriminatory" assessment may be politically acceptable but professionally useless, unless it reveals valid information about intelligent behavior for each group to which it is applied. Thus, for example, the pathetic attempt with SOMPA (System of Multicultural Pluralistic Assessment) is almost humorous. It is a hodgepodge of data which would take 50 IBM computers to unravel. The results offer no more to teaching than the I.Q. tests which it was designed to replace or augment. Indeed it even includes one of the I.Q. tests which its author had criticized earlier. Now SOMPA has joined mass production. The construct "adaptive behavior" has even less meaning than "intelligence."

If intelligence really exists in anything like the form which is represented in popular hypotheses, then the future may show something which has yet to be revealed. Varying patterns of thought - perhaps habits of thought would be more accurate - can be observed very readily today. However, the standardized "measurement" of mental "potential" or "ability" in either an absolute or a relative sense remains a hope and not a reality. After all the hocus-pocus, I.Q. testing and professional reasoning in education, using intelligence as a construct, tell us little more than a sensitive teacher already knows about a given child.

Educators bought the proverbial "pig-in-a-poke" when I.Q. testing and intelligence ideology was let into the tent. In doing so, they bought a new dependent with a ravenous appetite for resources. It is also a dependent that spends a great deal of time with its own cosmetics but no time at all helping with the housework. It has great fragrance but no substance. Maybe a diet or a fast would help.

Summary:

The standardized I.Q. tests which are in use in the schools are scientifically and pedagogically without merit. The construct "intelligence" is a hypothetical notion whose valid expression has yet to be born. I.Q. tests and the construct of intelligence can be discarded at present, and teaching strategies would be unaffected. To successful teachers the tests are at best a pure nuisance and at worst a reactive influence on teaching and learning. The tests are not simply culturally biased. That bias is only a symptom of the problem which is their scientific inadequacy. To say that "they are the best we have," is not to say that they contribute anything useful at all to instruction. The construct "intelligence" is embryonic and has heuristic value for research. Its utility for instruction remains to be demonstrated. School teachers and students should be relieved of the burden of this bad science and psychological ideology. Testmakers should come again when this product can help to make education better.

Selected Bibliography and References

Chomsky, Noam., Syntastic Structure. The Hague: Mouton 1957

*

Cole, M. & Scribner S., Culture and Thought. New York: John Wiley & Sons, 1974.

*

Elkind, David, A Sympathetic Understanding of the Child Six to Sixteen. Boston: Allyn and Bacon, 1971

Freire, Paulo, Education for Critical Consciousness. New York: Seabury, 1973.

Fuller, Renee, In Search of the I.Q. Correlation: A Scientific Whodunnit. Stony Brook, New York: Ball Stick Bird, 1977.

Furth, Hans G., Intellectual Health in School: What Can Piaget Contribute? In Piagetian Theory and its Implications for the Helping Professions. Los Angeles: University of Southern California, 1972, pp. 1-12.

Gallager, B. C., NAACP Report on Minority Testing. New York: NAACP Special Contribution Fund, May 1976.

Guthrie, Robert, Even the Rat Was White. New York: Harper and Row, 1976.

*

Hilliard, Asa G., III, Alternatives to I.Q. Testing: An Approach to the Identification of Gifted "Minority" Children, ERIC Clearinghouse on Early Childhood Education, ED, 1976.

Hilliard, Asa G., III, "Standardization and Cultural Bias: Impediments to the Scientific Study of Intelligence," Journal of Research and Development in Education, 12, 2, 1979, 47-58.

Hilliard, Asa G., III, "Standardized Testing and African-Americans," in Tyler, Ralph and Sheldon White. Teaching, Testing and Learning, Report of a Conference on Testing, National Institute of Education. Cape Cod, Mass., August 1978.

Hilliard, Asa G., III, "The Strengths and Weaknesses of Cognitive Tests for Young Children" in J.D. Andrews (Ed.). One Child Indivisible, Washington, D.C. The National Association for the Education of Young Children, 1976.

Hilliard, Thomas O., "The Role of Political and Social Action in the Prevention of Psychopathology of Blacks: A Mental Health Strategy for Oppressed People." University of Vermont, Fifth Conference on the Primary Prevention of Psychopathology, 1979.

Houts, Paul, The Myth of Measurability. New York: Hart, 1977.

Jennings, Wayne and Joe Nathan, "Startling/Disturbing Research on School Program Effectiveness." Phi Delta Kappan, March 1977, 568-572.

Jensen, Arthur, in "Jensen Condemns Minimum Competency, I.Q. Tests in the Schools," Phi Delta Kappan, 61, 1, 1979, p.75.

Johntz, William, Director Project SEED, Berkeley, California, 1976, (Interview).

Kamin, Leon, The Science and Politics of I.Q. New York: John Wiley & Sons, 1974.

Lawler, James M., I.Q. Heritability and Racism, New York: International Publishers, 1978.

Levi-Strauss, Claude, The Savage Mind, University of Chicago, 1966.

Levine, Murray, "The Academic Achievement Test." American Psychologist, 31, 1976, 228-238.

Maxwell, William, I.Q. Measuring and Monitoring Some Principles and Applications. Unpublished Papers, University of the South Pacific, 1979.

McClelland, David C., "Testing for Competence Rather Than for Intelligence." American Psychologist, 28, 1, 1973, 1-14

Nader, Ralph. "The Power and the Story of the Educational Testing Service." The Advocate George Washington University, Feb. 7, 1979, p. 9.

Orasanu, Judith, Rapymond P. McDermott, A. Wade Boykin, and the Laboratory of Comparative Human Cognition. "A Critique of Test Standardization," Social Policy, Sept./Oct. 1977, 8, 2, 61-67.

Smitherman, Geneva, Talking and Testifying: The Language of Black America. Boston: Houghton Mifflin, 1977.

Tyler, Ralph and Sheldon White, Testing, Teaching and Learning. Report of a Conference on Research on Teaching, Natl. Institute of Education, Cape Cod, Massachusetts Conference, August, 1978.

---

* Chase, Allan, The Legacy of Malthus: The Social Costs of the New Scientific Racism. New York: Alfred Knoph, 1977.

** Dreyfus, Joel & Charles Lawrence, III, The Baacke Case: The Politics of Inequality. New York: Harcourt, Brace, Jovanovich, 1979.

*** Hearnshaw, L.S., Cyril Burt Psychologist, Ithaca, New York: Cornell, 1979.

**** Tryon, Warren W. "The-Test-Trait-Fallacy" American Psychologist 34.5 1979, pages 402 to 406